

Inventors: Brian Searle, Surendra Dasari, Srinivasa Nagalla, Mark Turner

METHODS AND SYSTEMS FOR IDENTIFICATION OF MACROMOLECULES

ABSTRACT

A method is provided for identifying sequences of molecules and sequence modifications from mass spectrometry data. At least one *de novo* sequence is produced from mass spectrometry data of sequences of molecules,. At least one mass-based alignment is calculated between each *de novo* sequence and sequences in a sequence database. The molecular masses of molecules in the *de novo* sequence are compared to molecular masses of molecules in each sequence in the sequence database. Mass differences of modification sites are interpreted between the sequence in the sequence database and the *de novo* sequence that have been identified by the mass-based alignment as modifications identified in a modification catalog. At least one match score for the mass-based alignment is calculated that provides an indication of matching between the sequence in the sequence database and the *de novo* sequence. Sequences in the sequence database are identified from mass-based alignments in response to the match scores. Identifications of sequences in the sequence database are grouped from at least one *de novo* sequence into an identified macromolecule list that agrees with the *de novo* sequencing results.